

Link Formation on Twitter: The Role of Achieved Status and Value Homophily

Shujing Sun, Huaxia Rui

Simon Business School, University of Rochester

shujing.sun@simon.rochester.edu, huaxia.rui@simon.rochester.edu

Abstract—Homophily has been a widely recognized dominant factor in offline social network connection, which refers to one's propensity to seek interactions with others of similar status or values. Existing studies regarding homophily factors have been limited mostly to offline sociodemographic characteristics, such as race, gender, religion, education and occupation, which may not necessarily manifest homophily in online social network. Some researchers dabble in online social network, but they extract homophily characteristics from static user profile or link data, which has not incorporated the dynamic process of social network. To better understand the key factors in the establishment of online relationship, we explore a large data set on Twitter, which contains all initiated links by 1453 organizational Twitter users over three months. An initiated link refers to organization following a user who is currently not a follower of the organization. We crawl data on a daily basis and monitor whether the initiated one-way link ends up with a two-way relationship. Based on the established homophily theory, we define two online homophily factors: achieved status homophily (estimated by the gap of the followers count), value homophily (measured by the overlap ratio of common followee, Pearson correlation, and Cosine similarity between two users' tweets, respectively). We find that both homophily factors play a key role in the formation of online reciprocal relationship, and the effect of status homophily is larger for superior followee (one who has more followers than the corresponding organization) than for inferior followee (one who has less followers than the corresponding organization). Our finding not only extends the offline "individual-individual" homophily theory to the new online "organization-individual" relationship, but also provides Twitter users insight into extending their social network by strategically targeting followee.

Index Terms—Online Homophily; Social Media; Reciprocal Relationship Formation

I. Introduction

"Birds of a feather flock together", and "Homophily" is a formal term that captures the essence of this proverb, which refers to the tendency of individuals to associate and bond with similar others. The sociologists McPherson et al. point out "similarity breeds connection" in their classic 2001 paper [19]. They define "homophily" as the phenomenon that connection between similar people occurs at a higher rate than among dissimilar ones, and it structures network ties of every type.

Since then, many researchers have examined homophily of various dimensions in the offline social network, most of which focus on individual's attributes such as race, age, religion, education, occupation, gender, etc. For example, Kossinets and Watts investigate the role of homophily with respect to individual's occupation, gender and age on the formation of triadic closure [12]; Currarini et al. find that U.S. high school individuals tend to form friendship with

people of the same race [3]; Hegde et al. suggest that the U.S. venture capitalists are more likely to select start-ups with co-ethnic executives as partners for investment [7]; Louch shows that ascribed homophily (e.g., sex or race) and choice homophily (e.g., religion) can improve the likelihood of dyads relationship in personal network [17]; and Ruef et al. find that organizational founders in the U.S. are highly homogeneous by gender, ethnicity and status [26].

Meanwhile, some researchers have dabbled in studying the role of homophily in the realm of online social network. Their studies can be divided into two streams. The first stream concentrates on investigating the effect of offline homophily characteristics in the online setting, in which the offline characteristics are based on individual's self-reported attributes extracted through survey. For example, Thelwall and Mike show that except for gender homophily, other traditional sources of homophily (i.e. ethnicity, religion, age, country, marital status, etc.) are still important for friendship to form in MySpace [30]. Huang et al. verify the effect of homophily in age and game age on building collaborative relations in online games [8]. De Choudhury et al. [4] and Takhteyev et al. [29] examine the influence of geographic distance on Twitter. The second stream extracts homophily factors from the online social network, with a main focus on investigating value homophily in terms of topical interest. These studies focus on either static measures (e.g., tag [33], sentiment label [6], group membership [24]) that have not captured the dynamic change of value homophily, or the self-reported interest and user-defined profile that do not necessarily demonstrate the true-self in the online world ([14], [2]).

Due to the convenience of mobile and web-based technology, social media has greatly facilitated the development of online social network. Different from the traditional offline network, social media changes the way of communication, as well as the relationship between organizations and individuals. Firstly, social media provides both organizations and individuals with an equal opportunity to create an influential online presence easily and economically. With the unprecedented openness on social media sites, an ordinary individual can easily impair/enhance a long-established brands with a single negative/positive post [31]. Secondly, social media encourages contributions and feedback from everyone, which requires the organizations to actively participate in the discussions with their audience, instead of just disseminating information about itself. As Schroeder [27] points out, organizations need to

develop and maintain not just a brand but an online personality which is likable and well-respected, and with which individuals can develop a real sense of familiarity and emotional connection. In sum, with the power shifted to the general public and the requirements of two-way communication on social media platforms, the boundary between organizations and individuals are blurred, and organizations transform to personalized “pseudo-individual”.

Based on the above discussion, several questions occur to us: Can the classical homophily theory based on “individual-individual” relationship hold for the new “organization-individual” relationship in online social network? Besides the widely studied offline homophily, is there any online-specific homophily factor that truly captures such new type of link formation on social media? How can organizations (such as corporations, governments, not-for-profit corporations, educational institutions, etc.) make full use of their new “pseudo-individual” avatar to better connect with their audience and gain a larger follower base?

To attempt to answer the above questions, we need to first empirically distinguish homophily from influence. Since the difficulty lies in the limitation of static data, our paper uses a unique data set where we observe the dynamic link formation process. This allows us analyze the effect of homophily prior to the establishment of connection, and therefore to rule out influence as an alternative explanation for any association in the data. Specifically, we analyze the effect of online homophily on the formation of two-way relationship from an initiated one-way link using Twitter data. We distinguish two types of homophily: achieved status homophily, value homophily [15], and investigate how different factors affect one’s decision to respond to a following request from someone outside his/her circle of acquaintances. Different from previous research, we define and measure “online homophily” using variables extracted from the dynamic Twitter data. Using the logistic regression model with clustered standard errors, and controlling for a series of confounding factors, we find that both status and value homophily factors are important for relationship formation on social media platforms. Using the Conditional Logistic Regression model and the Logistic Regression model with Rare Events as the robustness check, our results still hold. In addition, we observe several interesting user-specific characteristics that influence the development of two-way relationship from one-way link, which sheds new light on strategically targeting followee for organizational Twitter users.

II. Theory and hypothesis

We develop our hypotheses based on the established definition proposed by Lazarsfeld, Merton, et al. Specifically, we distinguish two types of homophily: *status homophily*, in which similarity is based on the major formal or informal social status characteristics; and *value homophily*, which is based on values, attitudes, and beliefs [15]. *Status homophily* is further divided into two types: society-ascribed characteristics such as race, ethnicity, sex, or age, and achieved characteristics

like religion, education, occupation, or behavior patterns. *Value homophily* includes the wide variety of internal states that are presumed to shape our orientation toward future behavior, which is hard to measure in the offline world.

Since the self-reported offline characteristics may not be reliable (because a user might intentionally build a different image that is not true to himself/herself), and is not quite relevant to our context of studying “organization-individual” homophily, we focus on achieved status homophily and construct a new measure based on the readily available information on Twitter. We also overcome the difficulty in measuring value homophily by text mining users’ tweets and analyzing their link formation process. Next we will discuss each type of homophily and the corresponding hypotheses in detail.

A. Achieved Status Homophily — Vertical Measure

Status homophily has been extensively studied in offline social network. McPherson et al. suggest that both society-ascribed status and achieved status affect the formation of ties, but with different levels of effect [19]. Using data on 304 face-to-face groups from 10 communities, McPherson and Smith-Lovin find that friends are more similar on status dimensions (measured by years of education, age, occupational prestige, and sex) than chance [18]. Ruef, Aldrich, and Carter find that organizational founders in the U.S. are highly homogeneous in status measured by gender, ethnicity and occupational rank [26]. By analyzing a dynamic e-mail network at a large university, Kossinets and Watts find the effect of status homophily (categorized by undergraduate, graduate student, faculty, or staff) in the formation of triadic closure. For example, two students connected through a professor are less likely to form a direct tie than two students connected through another student, *ceteris paribus* [12]. Goodreau et al. suggest that the greater the difference in grade, the less likely students are to become friends [5]. Currarini et al. develop a model of friendship formation that explains the segregation pattern observed in social and economic network, and find that individuals tend to form friendship with people of the same race using a sample of U.S. high school students [3].

As we have discussed in the Introduction, organization transforms to a personalized “pseudo-individual” on social media platforms. Therefore, we expect that status homophily will affect the formation of such new “organization-individual” relationship. However, since offline status dimensions may not necessarily manifest homophily in online social network, we use the number of followers as the proxy for one’s achieved status on Twitter. This is because followers are Twitter’s most basic currency, which serve as the basis for one to build up social ties and get access to different resources. Quercia et al. [23], Romero et al. [25], and Hutto et al. [9] also suggest that the number of followers is a crucial indicator for one’s popularity and prestige. It is clear that offline status homophily shows significant effect in various kinds of “individual-individual” connections, we expect that online achieved status homophily in terms of the gap in followers count plays a similar role in online

social network. Hence, we develop the following Hypothesis 1.

Hypothesis 1: When *A* initiates a link to *B*, the higher the level of achieved status homophily (i.e. the lower the absolute difference in followers count), the more likely that *B* will follow back *A*.

In the context of this study, we focus on how a followee responds to an organization's initiated link. Since an individual can be either superior (i.e., followee has more followers than organization's Twitter account), or inferior (i.e., followee has fewer followers than organization's Twitter account) to the corresponding organization, the status homophily measured by the absolute difference in followers count could not reflect such asymmetry. Therefore, we further distinguish two types of achieved status homophily from the standpoint of followee, and we expect users with different vertical status to behave differently.

Steinfeld, Ellison and Lampe suggest that using social networking sites to form and maintain relationships is necessary for the accumulation of social capital, where social capital refers to the benefits one receives from his/her social relationships [28]. Many sociologists define social capital as resources embedded in one's social network, resources that can be accessed or mobilized through ties in the network [16]. For example, Putnam believes that social capital can be measured by the amount of trust and reciprocity in a community or between individuals [22]. Because users with more followers on Twitter can benefit more from the resources made available through their social relations, they will gain higher social capital accordingly. Kadushin points out that the desire for rank or status is a strong motive to form relationship on social network [10]. With the aim of social climbing, individuals tend to connect with or even defer to authority (those who are at the top of the occupational- or economic-class pyramid) in the offline world. We expect this tendency to be true in online social network. Specifically, users with a lower rank (in terms of the number of followers) have the incentive to connect with those with a higher rank. In addition, competence is a factor that is believed to influence the likelihood of becoming friends with another person. Parker and Seal define competence as how capable or able an individual is [20]. They suggest that people are more willing to develop committed relationship, such as friendships with those who are socially competent. Accordingly, followee who has fewer followers than an organization will find the organization more appealing, and therefore might have higher incentive to accept the organization's following request than followee with more followers.

Based on the above argument, we further define two types of achieved status homophily: *statusAbove* and *statusBelow*, which respectively indicates how much higher/lower a followee's status is than that of the organization. We expect that the effect of achieved status homophily on the inferior followee will be attenuated through the three mechanisms mentioned above. First, based on the social capital theory, an inferior followee has the incentive to follow back the

organization's account. This is because he/she can benefit from going beyond local circles and forging bridges to wider universes. Second, one tends to connect with a user of higher rank due to the incentive for social climbing. Third, the competence theory suggests that inferior followee is more attracted to those who possess social competence. Therefore, we propose Hypothesis 2, and expect that an inferior followee will be more likely to accept an organization's following request compared with the superior followee, *ceteris paribus*.

Hypothesis 2: When *A* initiates a link to *B*, the effect of the achieved status homophily will be attenuated if *B* has a lower status than *A*. Specifically, the higher the level of achieved status homophily (i.e. the lower the absolute difference in followers count), the higher the probability that *B* will follow back *A*. However, given the same level of achieved status homophily, *B*'s probability to follow back will be higher if *B* is inferior to *A* than if *B* is superior to *A*, *ceteris paribus*.

B. Value Homophily — Horizontal Measure

Values, attitudes, and beliefs influence one's behavior, interpersonal interactions, and even the hobbies one enjoys [13]. Lazarsfeld et al. define value homophily as the wide variety of internal states that are presumed to shape our orientation toward future behavior [15]. Gu et al. point out that homophily is rooted in the desire to feel confident about the accuracy of one's opinions and attitudes [6]. Or more generally, when we encounter others with similar views to ours, we feel good because the confirmation boosts confidence in the accuracy of our views.

In the existing homophily literature, there is no effective way of estimating value homophily, because one's belief, taste, and attitude are hidden attributes that are hard to observe. However, the emergence of social media platforms provides us with abundant resources. For example, Gu et al. [6] employ the sentiment labels collected from the Yahoo! Finance message boards to construct the opinion difference (i.e. the absolute difference of the sentiment value between the individual and the average sentiment of all postings) as a measure of value homophily. Zeng et al. [33] characterize the similarity of photos posted on Flickr according to the tags assigned to them. Rad and Benyoucef [24] use the common group memberships of users in the YouTube social network to measure the similarities between them. Aiello et al. [1] also observe a strong correlation between the social connectivity and the tagging/group participation behavior on social media platforms. Other researchers extract value homophily from the self-reported interest or user-defined tags on one's static profile ([14], [2]). As we can see, most of the existing studies merely focus on static measures (e.g., tag, sentiment label) that have not captured the dynamic change of value homophily, or user-defined characteristics that might not demonstrate homophily in the online world.

Given that the value homophily characteristics (e.g., shared values, attitudes, beliefs) are the same for any two actors in the social network, we define value homophily

as the horizontal homophily. This is to say that measures of value homophily is symmetric, regardless of the relative positions for the actors. Since following someone is similar to subscribing to a newspaper or magazine, which can reflect one's own choices and interests (a user can read all tweets from his/her followee through the one-way link), we define the value homophily as the ratio of mutual followee count to this user's total followee count. For example, given two users a and b , let A, B denotes the list of followee for a and b . Then the ratio of mutual followee that user b has with a is calculated by $\frac{|A \cap B|}{|B|}$, where $|B|$ and $|A \cap B|$ indicate the count of the corresponding followee list. We also try the measure $\frac{|A \cap B|}{|A \cup B|}$, and the result does not show much difference.

In order to directly capture the common interests of two Twitter users, we also construct two measures based on content analysis of the user's tweets. Specifically, we choose up to 400 tweets for each user before an organization initiates the link, and calculate the Pearson correlation and Cosine similarity based on latent semantic analysis for each "organization-individual" pair. If the "organization-individual" pair is very similar in their tweet content, we expect that the individual (i.e., followee) is more willing to follow the organization back, through which he/she can conveniently access the organization's tweets. Accordingly, we develop the following Hypothesis 3.

Hypothesis 3: When A initiates a link to B , the higher the level of value homophily (measured by ratio of mutual followee, Pearson correlation, or Cosine similarity) B has with A , the more likely that B will follow back A .

III. Data

A. Data Description

We choose charitable organizations to study the formation of two-way "organization-individual" relationship due to the following reasons. First, unlike profitable organizations, charities can not deliberately entice followee to follow back by providing coupons or promotional benefits. Second, unlike personal account, charities can not bridge relationship using the offline connection to a large extent. Third, social media is a promising platform for charities to conduct low-cost, highly effective campaign, in which choice of followee serves as a key step in extending their follower base, and future donors accordingly. Therefore, studying charity's initiated link not only provides us with an ideal set up to investigate the formation of reciprocal relationship from one-way link on social media, but also is beneficial for our society.

Ten brands are chosen in the study. These brands are chosen based on their top ranking on Forbes charity list, high Twitter adoption rate, and coverage of different industries. We take 1453 charity Twitter accounts into consideration, all of which are listed either on charity's national or branch official website. We use Twitter API to collect charity's tweets, follower list, followee list, and all related user's Twitter profile on a daily basis from February 21, 2016 to May 31, 2016. Table I lists the name of the ten brands, the corresponding industry they

belong to (from Charity Navigator), and the total number of Twitter accounts they have for national headquarter and local branches.

TABLE I. Basic information of the ten charity brands

Brand	Industry	Number of Branches
American Heart Association	Health	91
Easter Seals	Health	56
YMCA	Youth	446
Nature Conservancy	Environment/Animal	38
YWCA	Human Services	113
Catholic Charities USA	Human Services	73
American Red Cross	Human Services	39
United Way	Community Development	330
Habitat for Humanity	Community Development	99
Feeding America	Food, Agriculture and Nutrition	168

B. Summary Statistics

We define an initiated link as one such that a charity first follows someone who is not currently following the charity's Twitter account. The followee receives the following request, and then decides whether to follow back the charity or not. After removing invalid observations, there are 17,570 charity-initiated links, among which 1,406 links end up as two-way relationships by the end of May 31, 2016. Half of the Twitter users process the initiated link within 2 days, and 80% take less than 7 days to make the decision. Figure 1 is a heat map for the average follow back ratio in charity initiated link among different states within the U.S. It is clear that charities of different states differ in the success rate of bridging two-way relationship.

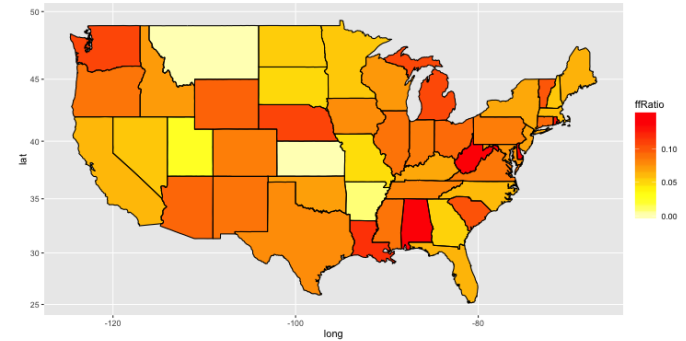


Fig. 1. Followee average follow back ratio by state

Table II lists the descriptions of the key variables used. For each individual followed by a charity, we assign them with a "follow back" dummy, where 1 indicates that the user followed by the charity also follows back the corresponding charity, and 0 otherwise. This is the dependent variable in the regression. The independent variables are classified into four groups: attributes of charity's Twitter profile (org_age , $org_favorites_count$, $org_followers_count$, $org_following_count$, org_list_count , $org_statuses_count$,

org_verified, *org_url*), attributes of followee’s Twitter profile (*age*, *favorites_count*, *followers_count*, *following_count*, *list_count*, *statuses_count*, *verified*, *Extrovert*), homophily factors (*statusDiff*, *statusAbove*, *statusBelow*, *followeeOverlap*, *Pearson*, *Cosine*), and control variables (*sameLocation*, *GoogleTrend*, *dow*, *BigFiveScore*).

We construct *statusDiff*, *statusAbove* and *statusBelow* based on user’s followers count, *followeeOverlap* based on the followee lists of the two parties associated in a link. For *Pearson* and *Cosine*, we adopt the latent semantic analysis (LSA) and compare the document to document similarity by analyzing up to 400 tweets for each of the two parties in a link. We use state-level location as proxy for geography homophily, because most of the charitable organizations in our data set are in charge of several regions simultaneously, so it is unlikely to locate a charity to a specific city or county. We extract followee’s location by mining their description, location and website link from their Twitter profile. We calculate the personality score for each Twitter user using their latest 400 tweets. Specifically, we first get the frequency of words in each category based on the LIWC 2001 dictionary [21], then we calculate the score of each personality trait using the correlation matrix between the Big Five personality traits and LIWC categories from Yarkoni’s paper [32].

TABLE II. Variable definition

Variable	Definition
follow back	Dummy variable indicating whether an individual follows back a charity or not
age	Number of days since the creation of a Twitter account
favorites_count	Number of tweets one has favorited in the account’s lifetime
followers_count	Number of followers a user currently has
following_count	Number of users an account is following, which is also referred as followee
list_count	Number of public lists that an user is a member of
statuses_count	Number of tweets&retweets issued by the user
verified	Dummy variable indicating if an account is verified
url	Dummy variable indicating if an account has an URL on its Twitter profile
Extrovert	Extrovert score of a followee, which is calculated by taking into consideration an user’s “url”, “profile image”, whether the account is “protected” and “geo-enabled”
statusDiff	$\log(followers_count - org_followers_count + 1)$
statusAbove	$\log(followers_count - org_followers_count + 1)$ if followee has more followers than charity; 0 otherwise
statusBelow	$\log(org_followers_count - followers_count + 1)$ if followee has fewer followers than charity; 0 otherwise
followeeOverlap	The ratio of mutual followee user U has with charity C , which is calculated by $\frac{ U \cap C }{ U }$
Pearson	Pearson product-moment correlation coefficient between the charity’s and followee’s tweets document.
Cosine	Cosine similarity between the charity’s and followee’s tweets document.
sameLocation	Dummy variable indicating whether a followee and a charity are located in the same state
GoogleTrend	One week lag Google Trend for each brand prior to each initiated link
dow	Day of the week for the date when charity initiates a link
BigFiveScore	Scores for each dimensions of the Big Five personality traits, which includes Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness

Table III shows the summary statistics of the key variables for charity and followee. Among the basic user profile vari-

ables, we can see that the range and standard deviation are very large, implying that our sample covers all possible pairs of organization and individual. Therefore, we are confident that our sample is representative enough for the study. We also check the correlation among the key independent variables, and find no collinearity issue.

TABLE III. Summary statistics

Vars.	Obs.	Mean	Std. Dev.	Min	Max
org_age	17570	1927	686	247	3282
org_favorites_count	17570	839	1941	0	22532
org_followers_count	17570	9859	120881	2	2736622
org_following_count	17570	1430	3755	1	42723
org_list_count	17570	162	950	0	17194
org_statuses_count	17570	3198	3865	1	29460
org_url	17570	0.97	0.16	0	1
org_verified	17570	0.02	0.14	0	1
age	17570	1852	811	7	3680
favorites_count	17570	1991	11441	0	440385
followers_count	17570	331487	2944482	1	77582588
following_count	17570	4332	52554	0	1573694
list_count	17570	1707	10284	0	214904
statuses_count	17570	9690	32170	0	1205710
verified	17570	0.18	0.38	0	1
Extrovert	17570	4.04	0.87	0	5
statusDiff	17570	8	3	0	18
statusAbove	17570	5	5	0	18
statusBelow	17570	3	4	0	15
followeeOverlap	17455	0.1	0.13	0	1
Pearson	17439	0.17	0.13	-0.79	0.87
Cosine	17439	0.26	0.13	0	0.88
sameLocation	11416	0.55	0.5	0	1
GoogleTrend	17570	78.54	13.31	21	100
follow back	17570	0.08	0.27	0	1

IV. Empirical Analysis and Results

A. Regression Model

In order to test the effect of online status and value homophily on the formation of two-way relationship given an initiated one-way link, we conduct a series of regressions. After log-transforming size-related measures, we have two sets of representative estimating equations as follows, in which the first regression model controls for charity’s brand and state fixed effect, while the second regression model controls for charity’s branch level fixed effect.

$$\begin{aligned}
\text{followBack}_{i,j} = & \alpha + \beta_0 \text{followeeProfile}_i + \beta_1 \text{charityProfile}_j \\
& + \text{statusHomophily}_{i,j} + \text{valueHomophily}_{i,j} \\
& + \text{geographyHomophily}_{i,j} + \text{GoogleTrend}_j \\
& + \text{BigFiveScore}_i + \text{dow} \\
& + \text{BrandFE} + \text{StateFE} + \epsilon_{i,j} \\
\text{followBack}_{i,j} = & \alpha + \beta_0 \text{followeeProfile}_i + \beta_1 \text{charityProfile}_j \\
& + \text{statusHomophily}_{i,j} + \text{valueHomophily}_{i,j} \\
& + \text{geographyHomophily}_{i,j} + \text{GoogleTrend}_j \\
& + \text{BigFiveScore}_i + \text{dow} + \text{BranchFE} + \epsilon_{i,j}
\end{aligned}$$

The dependent variable $\text{followBack}_{i,j}$ is a binary variable, indicating whether user i follows back charity j given that j sends a following request to i first (note that i has not followed

j before charity's initiated link). For independent variables, $followeeProfile_i$ includes variables extracted from followee i 's Twitter profile at the date when charity initiates a link, because these variables are the basis for the charity to make their following decision. $charityProfile_j$ includes variables extracted from charity j 's Twitter profile at the date when followee i follows charity j back, or the next day data after the date when charity sent the request if i did not follow j back (this is because the median time for one to follow back is two days). Details of $statusHomophily_{i,j}$ and $valueHomophily_{i,j}$ are defined in Table II. After verifying the significance of state, brand, and day of week fixed effect using Wald test in the baseline model, we include BrandFE (the brand that a branch belongs to), StateFE (the state where a charity locates), and dow (day of the week for the date when charity initiates a link) as control for all regressions. In order to control for the heterogeneity within each brand, we also include BranchFE in the second regression model.

McPherson et al. suggest that geography is a major source of homophily, and people are more likely to contact with those who are closer to them in geographic location than those who are distant [19]. Takhteyev et al. [29] and Huang et al. [8] show that spatial proximity still matters in online social network. Therefore, we also include $geographyHomophily$ to test whether offline physical distance still affects online social network connection. Since a branch is usually in charge of several regions within a state, it is hard to locate them to a specific city or county. So we use state level location as a proxy for $geographyHomophily$. Specifically, we use a dummy variable $sameLocation$ to denote whether i and j locate in the same state.

To check the robustness of the results, we also control for the weekly Google Trend for each of the ten brands in the full model. The reason why we do this is as follows. On the one hand, Google Trend reflects the searching frequency for a brand. For example, when someone receives the notification that "United Way" follows him/her on Twitter, he/she might search the brand for more details on Google. Therefore, Google Trend of a brand might directly affect one's perception and furthermore one's probability to follow back the charity. On the other hand, Google Trend indicates the popularity of an organization's online presence, which may influence the visiting frequency of the organization's Twitter profile. Accordingly, this will affect the number of followers, and further the achieved status homophily factor in our regression. Therefore, incorporating Google Trend not only improves the estimation efficiency, but also avoids any potential bias caused by the endogeneity.

Since the offline characteristics might also affect the followee's intention to follow back the corresponding charity, we include $BigFiveScore_i$ (see Table II for definition). Employing the LIWC 2001 dictionary [21], and the correlation matrix between the Big Five personality traits and LIWC categories [32], we calculate the personality score using up to 400 tweets for each followee prior to the date when an organization initiates a link. Unlike the easily manipulated user profile, one

can hardly disguise himself/herself by posting many tweets that are not true to himself/herself. We believe this is a good measure to represent followee's offline attribute.

We conduct a set of regressions using logistic and conditional logistic model, respectively. The logistic model gives us an overview of how online homophily factors affect followee's probability to follow back the charity. The conditional logistic model is to control charity branch-level fixed effect, and it differs from regular logistic regression in that the data are grouped by charity's twitter account and the likelihood is calculated relative to each group. To avoid the potential bias caused by the interdependence of observations belonging to the same charity brand, we cluster the data by brand and report clustered standard errors in the result.

B. Result and Analysis

The logistic and conditional logistic regression results are shown in Table IV and Table V, respectively. Clustered standard errors are reported in the parentheses. In both model, column 1 includes only the effect of status homophily measured by $statusDiff$. Column 2 incorporates two types of achieved status homophily: $statusAbove$ (representing how much higher a followee's status is than charity's status) and $statusBelow$ (representing how much lower a followee's status is than charity's status). Column 3–5 adds value homophily measured by $followeeOverlap$, $Pearson$, and $Cosine$, respectively. Column 6–8 adds geography homophily variable $sameLocation$ and weekly $GoogleTrend$ of each brand to test the robustness of the regression model. Table IV controls $BrandFE$, $StateFE$, dow fixed effect, and the $BigFiveScore$. Table V incorporates $BranchFE$ and the $BigFiveScore$.

In the logistic regression model, column 1 shows that the coefficient estimate of $statusDiff$ is significantly negative ($\beta = -0.1378$, $p\text{-value} < 0.01$). The result suggests that the larger the gap of online achieved status between charity and followee, the less likely that the followee will follow back the charity. This verifies our Hypothesis 1: A higher level of online status homophily results in the increased probability of online friendship formation.

From column 2, Table IV, we can see that the coefficient estimates of $statusAbove$ and $statusBelow$ are both significantly negative. We further test that their coefficients are significantly different from each other. Particularly, $statusAbove$ exerts a larger influence than $statusBelow$. As we can see that one unit increase in $statusAbove$ for a superior will reduce the log odds of following back by 0.1551, while one unit increase in $statusBelow$ for an inferior will reduce the log odds of following back by only 0.0924. This verifies our Hypothesis 2: The effect of the achieved status homophily will be attenuated if the followee has a lower status than the corresponding charity. As we discussed above, when a followee is inferior to the charity in terms of the number of followers, on the one hand he/she is reluctant to respond to the link due to the disparity in status, on the other he/she is attracted to charity's superiority through three mechanisms (including the need to gain social capital, the incentive to connect with the higher

rank actor, and the competence theory). Therefore, the overall effect of status homophily weakens for the inferior group.

Columns 3 – 5 suggest the significant positive effect of value homophily for all the three measures. This verifies our Hypothesis 3, which states that people with similar values (e.g., attitudes, beliefs, interests) have higher propensity to interact with each other on online social network. In this case, value homophily does not subsume the effect of status homophily.

Columns 6 – 8 show the results with all independent variables. The number of observations separately drops to 11,152, 11,145, and 11,145, because some Twitter users do not provide any location information or have protected accounts from which their tweets are not available to the public. We can see that the effect of status and value homophily are still significant for all cases, while the coefficient estimate for *sameLocation* is only very significant for the latter two cases (columns 7 and 8). This suggests that geographic location no longer exerts much effect in online relationship formation as in the offline world.

Several control variables extracted from charity's and followee's Twitter profile also show strong impact on the formation of the reciprocal relationship. We can see that a followee's follow back probability increases as "favorites_count" or "following_count" increases. This makes sense because both variables reflect one's engagement and activity level on Twitter. A higher level of engagement naturally leads to a higher tendency to bridge new relationship. Similarly, "verified" and "age" reflect one's prestige and credibility on Twitter, and thus negatively affect followee's probability to follow back the charity. "Extrovert" only shows significant positive effect in the first two columns, but its effect has been subsumed in the full regression model, which indicates that although a complete user profile (URL, profile image, unprotected, geo-enabled) reflect one's tendency and enthusiasm to interact with others, it may not be a very good measure to reflect one's true extraversion in reality. This verifies our assumption that social media users may not necessarily express their true self in the basic user profile.

In order to control for the individual's offline heterogeneity, we have calculated the Big Five personality score (Agreeableness, Conscientiousness, Extraversion, Neuroticism, Openness) by analyzing followee and organizations' tweets, and incorporated the *BigFiveScore* in all regressions. Because Twitter users are free to post anything they want, tweets serve as a voluntary expression of the individuals themselves. Therefore, the personality score in our paper provides a perfect control for Twitter user's offline characteristics. Due to the page limit, we didn't report their coefficient estimates in the table.

C. Robustness Check

As a robustness check, we further control for the charity branch-level fixed effect. As shown in Table V, all of the key results for our online homophily factors still hold.

Since the general logistic model is known to suffer from small-sample bias, we also refer to King and Zeng's logistic model with rare events [11]. King and Zeng suggest that their logistic model with rare events is suitable for the cases when the number of observations is small (under a few thousand) and the events are rare (under 5% or so). Clearly, our sample does not satisfy any of the criterion (17,570 observations with an average follow back ratio of 8%). Therefore, our result will not be biased using the regular logistic model. Nonetheless, in order to avoid any potential bias, we adopt King and Zeng's Rare Events Logistic Regression model as a robustness check. The results still hold. Due to the page limit, we do not include the table in this paper. The regression result is available upon request.

V. Conclusion

With the increasing popularity of social media platforms, online social network becomes an important channel for information creation and information sharing. While the dominant role of offline homophily in relationship formation has been studied extensively, online homophily has not been systematically analyzed using dynamic social media data. Exploring a dynamic data set containing all initiated links and the reciprocal decisions from Twitter, we construct two types of "online homophily" factors, and investigate their respective effect on the formation of two-way relationship from one-way link. This paper contributes to the established homophily literature in the following aspects.

First, existing homophily studies are based on "individual-individual" relationship. Considering that organizations transform to "pseudo-individuals" on social media platform, we extend the established homophily theory to "organization-individual" relationship.

Second, we construct new measures that are more appropriate in the online setting. Because the offline ascribed status homophily (e.g., race, ethnicity, sex, age) is not applicable to our context of studying "organization-individual" relationship, we focus on examining the effect of the achieved status homophily. Considering the asymmetry of the achieved status homophily, we further split it into two types, and use the difference in followers count as an estimation. We also overcome the difficulty of estimating value homophily in the existing literature. In particular, we calculate the followee overlap ratio using daily data, which captures the dynamics of value homophily in the online social network. In order to directly and accurately capture the common interests, we construct two other value homophily factors (Pearson correlation, Cosine similarity) through latent semantic analysis of user's tweets. Since the number of followers, the list of followee, and the tweets are publicly available information, we believe that the two parties associated in a link can better learn about each other through these indicators than through the unavailable offline homophily characteristics.

Third, our results show that "Online Birds of A Feather Still Flock Together", but with more interesting findings that have not been studied in the previous literature. We find that users

with higher achieved status and value homophily are more likely to form reciprocal relationship. However, the level of effect for achieved status homophily depends on the type of followee. Particularly, status homophily has a larger impact on the superior group than the inferior group. In terms of the geography homophily, we do not find much evidence of its effect in this study. This suggest that Twitter users can connect with each other from portable devices anytime and anywhere, and thus they are less sensitive to the physical distance as in the offline world.

These findings have important theoretical and practical implications. From a theoretical standpoint, our study sheds light on the key role of homophily in online social network. Different from previous research, we observe the dynamic link formation on Twitter, construct homophily factor using data from online social network, and use charitable organizations to exclude possible confounding factors. The longitudinal Twitter data enables us to observe the dynamics of link formation process, and to easily rule out influence as an alternative explanation for any association in the data. From a practical standpoint, our findings provide organizations insight into extending their social network by strategically targeting followee. For example, by a brief look at whether someone has a self-provided image, URL, description, number of followers, following, favorites, etc., the organizational Twitter users can roughly estimate the likelihood to be followed back by initiating a link. Accordingly, this enables organizations to better extend their online social network, and to develop a potential customer base.

References

- [1] L. M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines, and F. Menczer. Friendship prediction and homophily in social media. *ACM Transactions on the Web (TWEB)*, 6(2):9, 2012.
- [2] H. Bisgin, N. Agarwal, and X. Xu. Investigating homophily in online social networks. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 533–536. IEEE, 2010.
- [3] S. Currarini, M. O. Jackson, and P. Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.
- [4] M. De Choudhury, H. Sundaram, A. John, D. D. Seligmann, and A. Kelliher. "birds of a feather": Does user homophily impact information diffusion in social media? *arXiv preprint arXiv:1006.1702*, 2010.
- [5] S. M. Goodreau, J. A. Kitts, and M. Morris. Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks*. *Demography*, 46(1):103–125, 2009.
- [6] B. Gu, P. Konana, R. Raghunathan, and H. M. Chen. Research note: the allure of homophily in social media: Evidence from investor responses on virtual communities. *Information Systems Research*, 25(3):604–617, 2014.
- [7] D. Hegde and J. Tumlinson. Does social proximity enhance business partnerships? theory and evidence from ethnicity's role in us venture capital. *Management Science*, 60(9):2355–2380, 2014.
- [8] Y. Huang, C. Shen, and N. S. Contractor. Distance matters: Exploring proximity and homophily in virtual world networks. *Decision Support Systems*, 55(4):969–977, 2013.
- [9] C. J. Hutto, S. Yardi, and E. Gilbert. A longitudinal study of follow predictors on twitter. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 821–830. ACM, 2013.
- [10] C. Kadushin. *Understanding social networks: Theories, concepts, and findings*. OUP USA, 2012.
- [11] G. King and L. Zeng. Logistic regression in rare events data. *Political analysis*, 9(2):137–163, 2001.
- [12] G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. *science*, 311(5757):88–90, 2006.
- [13] M. W. Kraus, P. K. Piff, and D. Keltner. Social class, sense of control, and social explanation. *Journal of personality and social psychology*, 97(6):992, 2009.
- [14] H. W. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas. Homophily in the digital world: A livejournal case study. *Internet Computing, IEEE*, 14(2):15–23, 2010.
- [15] P. F. Lazarsfeld, R. K. Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.
- [16] N. Lin. A network theory of social capital. *The handbook of social capital*, 50:69, 2008.
- [17] H. Louch. Personal network integration: transitivity and homophily in strong-tie relations. *Social networks*, 22(1):45–64, 2000.
- [18] J. M. McPherson and L. Smith-Lovin. Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American sociological review*, pages 370–379, 1987.
- [19] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.
- [20] J. G. Parker and J. Seal. Forming, losing, renewing, and replacing friendships: Applying temporal parameters to the assessment of children's friendship experiences. *Child Development*, 67(5):2248–2268, 1996.
- [21] J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001, 2001.
- [22] R. D. Putnam. *Bowling alone: The collapse and revival of American community*. Simon and Schuster, 2001.
- [23] D. Quercia, J. Ellis, L. Capra, and J. Crowcroft. In the mood for being influential on twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, pages 307–314. IEEE, 2011.
- [24] A. A. Rad and M. Benyoucef. Similarity and ties in social networks: A study of the youtube social network. In *Proceedings of the Conference for Information Systems Applied Research ISSN*, volume 2167, page 1508, 2013.
- [25] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman. Influence and passivity in social media. In *Machine learning and knowledge discovery in databases*, pages 18–33. Springer, 2011.
- [26] M. Ruef, H. E. Aldrich, and N. M. Carter. The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *American sociological review*, pages 195–222, 2003.
- [27] H. Schroeder. The art of business relationships through social media. *IVEY Business Journal*.
- [28] C. Steinfield, N. B. Ellison, and C. Lampe. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, 29(6):434–445, 2008.
- [29] Y. Takhteyev, A. Gruz, and B. Wellman. Geography of twitter networks. *Social networks*, 34(1):73–81, 2012.
- [30] M. Thelwall. Homophily in myspace. *Journal of the American Society for Information Science and Technology*, 60(2):219–231, 2009.
- [31] Wikipedia. United Breaks Guitars. https://en.wikipedia.org/wiki/United_Breaks_Guitars.
- [32] T. Yarkoni. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3):363–373, 2010.
- [33] X. Zeng and L. Wei. Social ties and user content generation: Evidence from flickr. *Information Systems Research*, 24(1):71–87, 2013.

TABLE IV. Logistic regression results with clustered standard error

	<i>Dependent variable: follow back</i>							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
statusDiff	−0.1378*** (0.0229)							
statusAbove		−0.1551*** (0.0239)	−0.1558*** (0.0249)	−0.1382*** (0.0239)	−0.1366*** (0.0240)	−0.1609*** (0.0315)	−0.1435*** (0.0309)	−0.1427*** (0.0306)
statusBelow		−0.0924*** (0.0250)	−0.0944*** (0.0261)	−0.0794*** (0.0254)	−0.0794*** (0.0254)	−0.1016*** (0.0207)	−0.0835*** (0.0195)	−0.0839*** (0.0192)
followeeOverlap			0.7401** (0.3170)			0.7894** (0.3416)		
Pearson				1.3307*** (0.2870)			1.2901*** (0.3869)	
Cosine					1.4310*** (0.2989)			1.3523*** (0.4007)
sameLocation						0.1504* (0.0846)	0.2800*** (0.0909)	0.2728*** (0.0895)
GoogleTrend						0.0023 (0.0047)	0.0015 (0.0046)	0.0015 (0.0046)
favorites_count	0.0532** (0.0212)	0.0557** (0.0218)	0.0605*** (0.0224)	0.0527** (0.0231)	0.0506** (0.0233)	0.0846*** (0.0228)	0.0770*** (0.0228)	0.0756*** (0.0224)
following_count	0.5975*** (0.0353)	0.6158*** (0.0345)	0.6413*** (0.0392)	0.6357*** (0.0361)	0.6352*** (0.0358)	0.6809*** (0.0400)	0.6720*** (0.0404)	0.6713*** (0.0405)
list_count	−0.3662*** (0.0392)	−0.3041*** (0.0398)	−0.3162*** (0.0411)	−0.3320*** (0.0426)	−0.3308*** (0.0427)	−0.3298*** (0.0427)	−0.3302*** (0.0434)	−0.3299*** (0.0438)
statuses_count	−0.0788*** (0.0296)	−0.0739** (0.0294)	−0.0666** (0.0302)	−0.0754** (0.0310)	−0.0753** (0.0310)	−0.1032*** (0.0364)	−0.1173*** (0.0389)	−0.1168*** (0.0384)
verified	−0.5675*** (0.1192)	−0.5471*** (0.1251)	−0.5658*** (0.1346)	−0.4838*** (0.1284)	−0.4922*** (0.1273)	−0.4885*** (0.1575)	−0.4261*** (0.1481)	−0.4324*** (0.1492)
age	−0.3451*** (0.0559)	−0.3479*** (0.0548)	−0.3321*** (0.0532)	−0.3529*** (0.0499)	−0.3563*** (0.0499)	−0.2868*** (0.0666)	−0.3075*** (0.0626)	−0.3084*** (0.0628)
Extrovert	0.1098** (0.0531)	0.1073** (0.0537)	0.1032* (0.0540)	0.0925* (0.0551)	0.0940* (0.0547)	0.0960 (0.0740)	0.0866 (0.0756)	0.0886 (0.0751)
org_favorites_count	0.0641** (0.0278)	0.0632** (0.0279)	0.0719*** (0.0274)	0.0709*** (0.0262)	0.0706*** (0.0264)	0.0426 (0.0290)	0.0416 (0.0294)	0.0417 (0.0296)
org_following_count	0.0352 (0.0925)	0.0163 (0.0916)	−0.0244 (0.0960)	−0.0055 (0.0852)	−0.0051 (0.0854)	0.0056 (0.0732)	0.0380 (0.0656)	0.0392 (0.0660)
org_list_count	0.0506 (0.0425)	−0.0167 (0.0455)	−0.0335 (0.0412)	−0.0478 (0.0382)	−0.0502 (0.0383)	−0.1020* (0.0569)	−0.1257** (0.0528)	−0.1277** (0.0531)
org_statuses_count	0.0829 (0.0807)	0.0773 (0.0850)	0.0926 (0.0814)	0.0839 (0.0769)	0.1011 (0.0734)	0.1288 (0.0844)	0.1166 (0.0818)	0.1318* (0.0799)
org_url	0.1777 (0.2610)	0.1554 (0.2605)	0.0046 (0.2235)	0.0365 (0.2001)	0.0222 (0.1972)	0.3186 (0.3615)	0.3229 (0.3420)	0.3117 (0.3410)
org_verified	0.0772 (0.3152)	0.0674 (0.3349)	0.0393 (0.3415)	0.0077 (0.3243)	0.0101 (0.3247)	0.4707 (0.3001)	0.4575 (0.2827)	0.4573 (0.2822)
org_age	−0.0581 (0.0864)	−0.0683 (0.0882)	−0.0393 (0.0792)	−0.0392 (0.0809)	−0.0392 (0.0792)	−0.0159 (0.0745)	−0.0094 (0.0750)	−0.0102 (0.0740)
brand, state, dow FE, BigFiveScore	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,570	17,570	17,455	17,439	17,439	11,152	11,145	11,145
R ²	0.1769	0.1793	0.1814	0.1846	0.1848	0.1823	0.1850	0.1850
χ ²	1,351.3480***	1,370.4400***	1,378.1100***	1,403.3650***	1,404.8500***	886.9936***	900.9989***	901.1293***

Note:

*p<0.1; **p<0.05; ***p<0.01

TABLE V. Conditional logistic regression results with clustered standard error

	<i>Dependent variable: follow back</i>							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
statusDiff	−0.1180*** (0.0219)							
statusAbove		−0.1307*** (0.0222)	−0.1343*** (0.0223)	−0.1158*** (0.0227)	−0.1144*** (0.0227)	−0.1379*** (0.0290)	−0.1174*** (0.0294)	−0.1160*** (0.0294)
statusBelow		−0.0651*** (0.0252)	−0.0721*** (0.0253)	−0.0580** (0.0255)	−0.0577** (0.0255)	−0.0746*** (0.0322)	−0.0551** (0.0325)	−0.0547** (0.0324)
followeeOverlap			0.8635*** (0.2834)			1.1615*** (0.3684)		
Pearson				1.5472*** (0.2682)			1.7798*** (0.3546)	
Cosine					1.6110*** (0.2832)			1.8245*** (0.3751)
sameLocation						0.1180 (0.0990)	0.2860*** (0.1033)	0.2741*** (0.1027)
GoogleTrend						−0.0016 (0.0064)	−0.0024 (0.0065)	−0.0025 (0.0065)
favorites_count	0.0400** (0.0160)	0.0415** (0.0161)	0.0479** (0.0163)	0.0372* (0.0163)	0.0346* (0.0163)	0.0760*** (0.0222)	0.0645*** (0.0221)	0.0618*** (0.0222)
following_count	0.4987*** (0.0281)	0.5201*** (0.0288)	0.5455*** (0.0307)	0.5321*** (0.0294)	0.5314*** (0.0294)	0.6059*** (0.0425)	0.5795*** (0.0403)	0.5775*** (0.0403)
list_count	−0.3180*** (0.0313)	−0.2572*** (0.0342)	−0.2654*** (0.0345)	−0.2796*** (0.0348)	−0.2776*** (0.0348)	−0.2872*** (0.0490)	−0.2843*** (0.0489)	−0.2840*** (0.0489)
statuses_count	−0.0509** (0.0267)	−0.0477* (0.0267)	−0.0413* (0.0269)	−0.0466* (0.0273)	−0.0462* (0.0274)	−0.0847*** (0.0372)	−0.0955*** (0.0378)	−0.0942*** (0.0379)
verified	−0.5452*** (0.1454)	−0.5015*** (0.1464)	−0.5213*** (0.1484)	−0.4520*** (0.1479)	−0.4657*** (0.1478)	−0.3967*** (0.1803)	−0.3277** (0.1800)	−0.3415*** (0.1798)
age	−0.2389*** (0.0382)	−0.2374*** (0.0381)	−0.2294*** (0.0385)	−0.2516*** (0.0387)	−0.2547*** (0.0387)	−0.1822*** (0.0521)	−0.2183*** (0.0524)	−0.2174*** (0.0523)
Extrovert	0.0896* (0.0370)	0.0893* (0.0369)	0.0878* (0.0370)	0.0812* (0.0372)	0.0838* (0.0371)	0.0719 (0.0514)	0.0659 (0.0515)	0.0698 (0.0514)
org_favorites_count	0.4637 (0.2396)	0.4661* (0.2391)	0.4718* (0.2386)	0.4944* (0.2384)	0.4901* (0.2389)	0.2748 (0.2704)	0.2761 (0.2706)	0.2776 (0.2710)
org_following_count	0.3900* (0.2750)	0.3706* (0.2733)	0.3792* (0.2731)	0.3404 (0.2734)	0.3459* (0.2731)	0.5100* (0.3206)	0.4654* (0.3202)	0.4770* (0.3199)
org_list_count	2.5559*** (0.6820)	2.4756*** (0.6782)	2.5193*** (0.6797)	2.4746*** (0.6777)	2.4577*** (0.6788)	1.8896** (0.7559)	1.8839** (0.7534)	1.8600** (0.7545)
org_statuses_count	−0.1328 (0.8079)	−0.0321 (0.8106)	−0.0813 (0.8117)	−0.1328 (0.8129)	−0.1333 (0.8130)	0.0698 (0.9371)	0.0636 (0.9360)	0.0470 (0.9367)
org_age	3.9636** (1.8125)	3.8821** (1.8097)	3.8211** (1.8133)	3.8550** (1.8170)	3.8652** (1.8162)	2.7211 (2.4517)	2.6542 (2.4613)	2.6133 (2.4622)
branch, dow FE, BigFiveScore	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,570	17,570	17,455	17,439	17,439	11,152	11,145	11,145
R ²	0.0504	0.0515	0.0519	0.0534	0.0533	0.0492	0.0507	0.0506
Max. Possible R ²	0.4212	0.4212	0.4215	0.4221	0.4221	0.3853	0.3860	0.3860
Log Likelihood	−4,349.4690	−4,339.2260	−4,311.9870	−4,302.9770	−4,303.3010	−2,432.4820	−2,427.7820	−2,428.4340
Wald Test	9,414.2300***	23,041.1800***	356,301.8000***	593,243.5000***	1,771,986.0000***	2,518.0200***	1,202.9900***	991.8500***
LR Test	908.9355***	929.4207***	930.8916***	956.2094***	955.5630***	562.7217***	579.6785***	578.3750***
Score (Logrank) Test	750.7430***	764.4425***	773.5006***	796.9185***	797.9576***	465.4799***	482.2968***	482.5589***

Note:

*p<0.1; **p<0.05; ***p<0.01